AN ALMOST-ALWAYS MEASURE OF THE FAILURE OF UNIQUE FACTORIZATION

PAUL POLLACK AND ENRIQUE TREVIÑO

Abstract. Let K be a number field. For each nonzero, nonunit α belonging to the ring of integers $\mathcal{O}_K$ of K, the **elasticity** $\rho(\alpha)$ of α is the largest possible ratio m/n, where m and n range over lengths of factorizations of α into irreducible elements of $\mathcal{O}_K$. A celebrated theorem of Narkiewicz, Steffan, and Valenza asserts that the largest value of $\rho(\alpha)$ admits a simple description in terms of the Davenport constant of the class group of K. Less well-known is that the values $\rho(\alpha)$ concentrate around a single real number $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ (as shown by Narkiewicz and Śliwa). We describe $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ in terms of a game played on the class group of K, and we analyze this game for two new families of groups.

1. Introduction

1.1. **Stretching: the truth about unique factorization.** Recall that when D is an integral domain, a nonzero nonunit $\pi \in D$ is called **irreducible** if π cannot be written as a product of two nonunits in D. We call D **atomic** if every nonzero nonunit of D admits at least one representation as a product of irreducibles of D. A **unique factorization domain** (or **factorial domain**, or **UFD**) is a domain where every nonzero nonunit factors *uniquely* into irreducibles. That is, D is atomic, and whenever two products of irreducibles coincide, say

$$\pi_1 \cdots \pi_k = \rho_1 \cdots \rho_\ell, \tag{1}$$

there is an obvious explanation for this: Namely,

(i) $k = \ell$, *and*

(ii) for some permutation τ of $\{1, 2, 3, \dots, k\}$, and some units $\varepsilon_1, \dots, \varepsilon_k \in D$, we have $\pi_{\tau(i)} = \varepsilon_i \rho_i$ for all $i = 1, 2, \dots, k$.

Many of the domains one encounters in a first algebra course are UFDs (including $\mathbb{Z}$, $\mathbb{F}[x]$ for $\mathbb{F}$ a field, $\mathbb{Z}[x]$, and $\mathbb{Z}[i]$), but as soon as one ventures out into the mathematical wild, UFDs start to appear thin on the ground.

If D is not a UFD, how far away is it from being one? Put slightly differently, when unique factorization breaks down, how disastrous is its failure?

In 1980, R. J. Valenza [17] — inspired by work of L. Carlitz twenty years prior [2] — introduced "elasticity" as a possible answer to these questions. Let D be an atomic domain. For each nonzero nonunit $\alpha \in D$, we define the **length spectrum** $\mathcal{L}(\alpha)$ of α as the set of all k for which α admits a factorization into precisely k irreducible elements of D:

$$\mathcal{L}(\alpha) = \{k : \alpha = \pi_1 \cdots \pi_k \text{ for some irreducibles } \pi_i\}.$$

2020 *Mathematics Subject Classification.* Primary 11R27; Secondary 13A05, 13F15.

The **elasticity** $\rho(\alpha)$ of α is defined as the "multiplicative diameter" of $\mathcal{L}(\alpha)$. More precisely,

$$\rho(\alpha) = \frac{\sup \mathcal{L}(\alpha)}{\inf \mathcal{L}(\alpha)} \in [1, \infty].$$

The elasticity of the domain D itself, denoted $\rho(D)$, is taken to be

$$\sup_{\alpha} \rho(\alpha),$$

where α ranges over all the nonzero nonunits of D.

If D is atomic (and not a field), then $\rho(D) \geq 1$. Having $\rho(D) = 1$ can be thought of as being halfway to unique factorization, in the following sense: Whenever two products of irreducibles coincide as in (1), $k = \ell$. That is, we have condition (i) (but not necessarily (ii) !) in the above definition of a UFD. Accordingly, domains with $\rho(D) = 1$ are called **half-factorial** (see [4] for a survey of research into this class of domains).

This article concerns elasticities of elements from rings of integers of number fields.

A **number field** is a field K containing $\mathbb{Q}$ for which $[K : \mathbb{Q}] < \infty$. To each number field K is associated a **ring of integers**, denoted $\mathcal{O}_K$, which stands in relation to K as the familiar ring $\mathbb{Z}$ stands in relation to $\mathbb{Q}$. (For precise definitions of all terms from algebraic number theory used here, see, e.g., [16].) The rings $\mathcal{O}_K$ are all atomic domains, but they are frequently non-UFDs. For example, the number field $K = \mathbb{Q}(\sqrt{-5})$ has ring of integers $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$, and the failure of unique factorization in $\mathbb{Z}[\sqrt{-5}]$ is well-known. Indeed, if you asked a random person on the street for a counterexample to unique factorization, 99% of them would back away slowly, but the remaining 1% would trot out the equation

$$(1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}) = 2 \cdot 3,$$

violating condition (ii) in our UFD definition.

While the nonzero, nonunit elements of the rings $\mathcal{O}_K$ need not factor uniquely into irreducible elements (up to order and unit-multiples), a foundational 19th century result of Richard Dedekind asserts that the nonzero, nonunit *ideals* of $\mathcal{O}_K$ always factor uniquely into irreducible ideals (up to the order of the factors). Here the product of two ideals I, J is defined as the smallest ideal containing all the products $\alpha\beta$, with $\alpha \in I$, $\beta \in J$, and an ideal is **irreducible** if it is not (0), not $(1) = \mathcal{O}_K$, and cannot be expressed as IJ for any nonunit ideals I and J.[1]

If α and β are elements of $\mathcal{O}_K$, the product of the principal ideals generated by α and β is the principal ideal generated by the product $\alpha\beta$. It follows that $\mathcal{O}_K$ is a UFD precisely when its principal ideals factor uniquely into principal ideals that are irreducible *as principal ideals*, meaning not (0) or (1) and not the product of two nonunit principal ideals. We have seen that this flavor of unique factorization is sometimes too much to hope for. In these cases, Dedekind tells us it is not $\mathcal{O}_K$ that is deficient, but our overly provincial vision. Restoring unique factorization is as simple as granting nonprincipal ideals full citizenship.

There is a gadget attached to every number field K that functions as a bridge between the factorization-friendly world where all ideals are considered and the more temperamental realm of elements (or equivalently, of only principal ideals): the **class group** $\mathrm{Cl}(K)$. Formally, $\mathrm{Cl}(K)$ is the quotient of the group of (nonzero, fractional) ideals of $\mathcal{O}_K$ by the group of (nonzero, fractional)

[1]Dedekind's result is usually stated in terms of *prime* ideals. In fact, in $\mathcal{O}_K$ the class of prime ideals and the class of irreducible ideals coincide, as a consequence of the maxim "to divide is to contain".

principal ideals. In informal contexts, number theorists often introduce the class group as a device that "measures the failure of unique factorization."

We now have two yardsticks with which to measure how far away $\mathcal{O}_K$ is from being a UFD: the elasticity $\rho(\mathcal{O}_K)$, which is a number (a priori possibly ∞), and $\mathrm{Cl}(K)$, a finite abelian group. The link between the two is provided by an extraordinarily elegant result, due to R. J. Valenza[2] [17], W. Narkiewicz [12], and J.-L. Steffan [15].

Before stating their theorem, we need a definition from additive combinatorics. If G is a finite abelian group, the **Davenport constant** of G, denoted $\mathrm{Dav}\, G$, is the smallest positive integer D possessing the following property:

Every sequence $g_1, \ldots, g_D \in G$ possesses a subsequence summing to 0 in G.

(In what follows, groups are always written additively, with the identity denoted by 0. "Subsequence" always means "nonempty subsequence".) The Davenport constant always exists, and in fact, $\mathrm{Dav}\, G \leq \#G$, with equality for cyclic groups G (we prove this as Lemma 5.1 below).

Theorem A. *Let K be a number field. If $\mathrm{Cl}(K)$ is trivial, then $\mathcal{O}_K$ is a UFD and $\rho(\mathcal{O}_K) = 1$. Otherwise,*

$$\rho(\mathcal{O}_K) = \frac{1}{2}\,\mathrm{Dav}\,\mathrm{Cl}(K).$$

As an example, consider $K = \mathbb{Q}(\sqrt{-5})$. The class group of $\mathbb{Q}(\sqrt{-5})$ is cyclic of order 2, so that Theorem A yields $\rho(\mathbb{Z}[\sqrt{-5}]) = 1$. That is, $\mathbb{Z}[\sqrt{-5}]$ is a half-factorial domain. On the other hand, the class group of $K = \mathbb{Q}(\sqrt{-26})$ is cyclic of order 3, and so its ring of integers $\mathbb{Z}[\sqrt{-26}]$ has elasticity $\frac{3}{2}$. That the elasticity is at least $\frac{3}{2}$ can be seen from the equation

$$(1 + \sqrt{-26})(1 - \sqrt{-26}) = 3 \cdot 3 \cdot 3.$$

(Of course, one should check that all the factors on both sides are irreducible. We leave this to the reader.)

We have set up our definitions so that for every nonzero nonunit $\alpha \in \mathcal{O}_K$, it is automatic that

$$1 \leq \rho(\alpha) \leq \rho(\mathcal{O}_K).$$

It is natural to wonder how the numbers $\rho(\alpha)$ are distributed between these two extremes as α varies. It does not seem to be so well-known that a satisfactory answer to this question was given by Narkiewicz and J. Śliwa [13] already in 1977 (yes, a few years before the formal study of elasticity!).

For each number field K and each property P pertaining to principal ideals of $\mathcal{O}_K$, we say that P holds for **almost all** principal ideals if it holds for asymptotically 100% of them, when ordered by norm. More precisely, we require that

$$\lim_{X \to \infty} \frac{\#\{\text{principal ideals } (\alpha) : |N\alpha| \leq X, \text{P holds for } (\alpha)\}}{\#\{\text{principal ideals } (\alpha) : |N\alpha| \leq X\}} = 1,$$

where $N\alpha$ denotes the norm of α.

[2]The 1990 publication date of Valenza's paper is somewhat misleading. Valenza's paper was received at the journal on November 19, 1980.

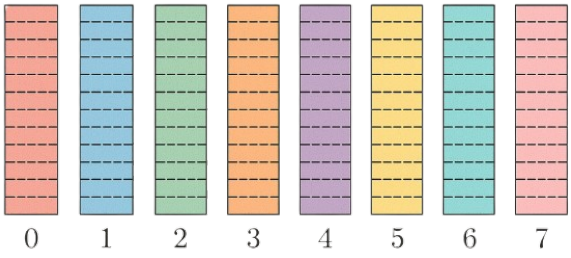

FIGURE 1. Initial configuration of $(\mathbb{Z}/8\mathbb{Z})$-solitaire with stacks of height $X = 8$.

We can now state the theorem of Narkiewicz and Śliwa, which asserts that the elasticities $\rho(\alpha)$ concentrate around a single value, here denoted $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ (`typ` for "typical").[3]

Theorem B. *Let K be a number field. There is a real number $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ for which the following holds. For each fixed $\epsilon > 0$, almost all principal ideals (α) of $\mathcal{O}_K$ are such that*

$$(1-\epsilon)\rho_{\mathrm{typ}}(\mathcal{O}_K) < \rho(\alpha) < (1+\epsilon)\rho_{\mathrm{typ}}(\mathcal{O}_K).$$

Furthermore, $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ depends only on (the isomorphism class of) $\mathrm{Cl}(K)$.

The way we have stated Theorem B comes off as a little coy in comparison with Theorem A. The elasticity $\rho(\mathcal{O}_K)$ is half the Davenport constant of $\mathrm{Cl}(K)$ (unless $\mathcal{O}_K$ is a UFD). *Which* function of $\mathrm{Cl}(K)$ is $\rho_{\mathrm{typ}}(\mathcal{O}_K)$? The proof in [13] supplies an answer, in terms of the solution to a certain linear programming problem. One of our primary motivations for writing is to offer an entirely equivalent but more whimsical description (Theorem B′ below), in terms of a constant associated to optimal play in a game we call **group solitaire**. We suspect this game will be of independent interest, and we hope that our advertisement here will encourage its further study.

1.2. **Playing by yourself, in a group: G-solitaire.** Let G be a finite abelian group, and let X be an even positive integer. The game of **G-solitaire, initialized at height X**, is played as follows: We begin with a "table" consisting of $\#G$ stacks of poker chips, labeled by the distinct elements of G, each stack having the same starting height (number of chips) X. We view the chips in each stack as copies of the corresponding element of G.

Since X is even, the sum of the chips in the initial configuration is equal to

$$X\sum_{g\in G} g = \frac{X}{2}\sum_{g\in G} g + \frac{X}{2}\sum_{g\in G} g = \frac{X}{2}\sum_{g\in G} g + \frac{X}{2}\sum_{g\in G}(-g) = 0. \tag{2}$$

A move in G-solitaire consists of discarding any (nonempty) collection of chips that sums to 0 in G, *as long as no proper subcollection also sums to* 0. In view of (2), any sequence of legal moves eventually clears the table. The objective in G-solitaire is to clear the table with the smallest possible number of moves, denoted $\Sigma(G; X)$.

For any G, the only way chips from the 0-stack can be removed is one-at-a-time. When G is trivial, every chip belongs to the 0-stack, and thus $\Sigma(G; X) = X$. Now suppose $G \cong \mathbb{Z}/2\mathbb{Z}$. No

[3]Narkiewicz and Śliwa do not state Theorem B directly: Their concern is with the individual maximum and minimum elements of $\mathcal{L}(\alpha)$ (rather than the ratio of the two). Furthermore, they restrict to rational integers α, rather than letting α range over $\mathcal{O}_K$. But their methods certainly suffice to establish Theorem B, and we feel it is appropriate to credit them.

matter how we clear the table, we spend X moves clearing the 0-stack, and the X chips in the 1-stack end up removed 2-at-a-time. Hence, $\Sigma(G;X) = X + \frac{1}{2}X = \frac{3}{2}X$.

In these two examples, there is only one way to clear the table (up to the order moves are made). But matters get more interesting already for $G \cong \mathbb{Z}/3\mathbb{Z}$. As always, we are forced to clear the 0-stack in X moves, one-at-a-time, but now there are multiple ways to finish off. For instance, we can clear the $2X$ nonzero chips 2-at-a-time, each move discarding 1 chip from the 1-pile and 1 chip from the 2-stack. This uses $X + X = 2X$ total moves. A more efficient procedure is to remove 3 chips from the 1-stack $\lfloor X/3 \rfloor$ times, 3 chips from the 2-stack $\lfloor X/3 \rfloor$ times, and then to sweep up what is left in at most 2 further moves. This latter strategy requires only $\approx X + \frac{2}{3}X = \frac{5}{3}X$ moves.

The following result (implicit in [13]) could be considered "The Fundamental Theorem of G-Solitaire." It guarantees that for each fixed G, the function $\Sigma(G;X)$ grows linearly with X with a well-defined constant of proportionality.

Proposition 1.1. *For each fixed finite abelian group G, there is a positive constant* $\operatorname{Clr} G$ *for which*

$$\lim_{X\to\infty} \frac{\Sigma(G;X)}{X} = \operatorname{Clr} G.$$

Furthermore, $\operatorname{Clr} G$ *is a rational number.*

We call $\operatorname{Clr} G$ the **clearing constant** associated to G. For instance, the trivial group has clearing constant 1, while $\operatorname{Clr} \mathbb{Z}/2\mathbb{Z} = \frac{3}{2}$ and $\operatorname{Clr} \mathbb{Z}/3\mathbb{Z} \le \frac{5}{3}$. Later we shall see that $\operatorname{Clr} \mathbb{Z}/3\mathbb{Z} = \frac{5}{3}$ (see eq. (8)).

We can now complete the statement of Narkiewicz and Śliwa's Theorem B.

Theorem B′. *For each number field K, Theorem B holds with*

$$\rho_{\text{typ}}(\mathcal{O}_K) = \frac{1 + \#\mathrm{Cl}(K)}{2 \operatorname{Clr} \mathrm{Cl}(K)}.$$

Shortly (§2) we will sketch the proofs of Proposition 1.1 and Theorem B′. We pause first for a few remarks.

From Theorem B′ and our above discussion of clearing constants, we see that $\rho_{\text{typ}}(\mathcal{O}_K) = 1$ whenever $\#\mathrm{Cl}(K) = 1$ or 2. This is no surprise: In these cases, $\rho(\mathcal{O}_K) = 1$ (by Theorem A), so that $\rho(\alpha) = 1$ for *all* nonzero, nonunit $\alpha \in \mathcal{O}_K$. By contrast, when $\#\mathrm{Cl}(K) = 3$, we have $\rho(\mathcal{O}_K) = \frac{3}{2}$ while $\rho_{\text{typ}}(\mathcal{O}_K) = \frac{6}{5}$. (One can prove more generally that $1 < \rho_{\text{typ}}(\mathcal{O}_K) < \rho(\mathcal{O}_K)$ whenever $\mathcal{O}_K$ is not half-factorial.)

Of course, one would like to determine the values of $\operatorname{Clr} G$ for as wide a class of groups G as possible. The first efforts in this direction seem to have been made by S. Allen and P. A. B. Pleasants [1].[4] They determined $\operatorname{Clr} G$ for all homocyclic p-groups, meaning all groups of the form $(\mathbb{Z}/p^r\mathbb{Z})^s$, where p is prime and r, s are positive integers (we state their result later as eq. (11)). They also found, by ad hoc calculations, that

$$\operatorname{Clr} \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z} = \frac{8}{3} \quad \text{and} \quad \operatorname{Clr} \mathbb{Z}/6\mathbb{Z} = \frac{13}{6}.$$

[4]Their work takes place in the same context as that of Narkiewicz and Śliwa [13]; see footnote 3.

Our principal theorems place these last two results in a more general framework.

Theorem 1.2. *For every prime p,*

$$\operatorname{Clr} \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z} = 1 + \frac{2p^2 - p - 1}{2p - 1}.$$

(We write the constant in the form $1 + \cdots$ to isolate the uninteresting contribution from the 0-stack, which is always cleared one chip at a time.)

Theorem 1.3. *For every pair of distinct primes p and q,*

$$\operatorname{Clr} \mathbb{Z}/pq\mathbb{Z} = 3 - \frac{1}{p} - \frac{1}{q}.$$

Our proofs depend fundamentally on U. Krause's notion of the **cross number** of a finite abelian group [9]. The observation that the cross number of the class group carries information about factorizations in $\mathcal{O}_K$ is not novel by any means. Indeed, Krause introduced his invariant in order to establish a factorization-based characterization of number fields with class groups cyclic of prime power order! (See the start of §3, and cf. [10], [3].) However, as far as we are aware, this is the first use of the cross number to investigate "typical elasticities."

2. Laying the groundwork: Proof sketches for Proposition 1.1 and Theorem B

2.1. **Sketch of the proof of Proposition 1.1.** If G is a finite abelian group, we let $\mathbb{R}^G$ denote the $\mathbb{R}$-vector space of functions from G to $\mathbb{R}$, viewing the elements of $\mathbb{R}^G$ as vectors of real numbers indexed by G. For $\boldsymbol{v} \in \mathbb{R}^G$ and $g \in G$, the gth component of $\boldsymbol{v}$ will be denoted $\boldsymbol{v}[g]$. We can, and do, identify each move in G-solitaire with the element of $\mathbb{R}^G$ whose gth component counts the number of removed chips from the g-stack.

We let $\mathbb{L}_G \subseteq \mathbb{R}^G$ denote the collection of legal moves in G-solitaire. Equivalently, $\mathbb{L}_G$ is the set of $\boldsymbol{v} \in \mathbb{R}^G$ with nonnegative integer entries, not all zero, satisfying

(a) $\sum_{g \in G} \boldsymbol{v}[g]g = 0$ in G, *and*

(b) whenever $\boldsymbol{w}$ is dominated by $\boldsymbol{v}$, in the sense that $0 \le \boldsymbol{w}[g] \le \boldsymbol{v}[g]$ for each $g \in G$, and $\sum_{g \in G} \boldsymbol{w}[g]g = 0$, either $\boldsymbol{w} = \boldsymbol{0}$ (the zero vector) or $\boldsymbol{w} = \boldsymbol{v}$.

Since every legal move has length at most $\operatorname{Dav} G$, the set $\mathbb{L}_G$ is finite. List its elements as $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$. Define the vector $\mathbf{1} \in \mathbb{R}^G$ by setting $\mathbf{1}[g] = 1$ for all $g \in G$. Then

$$\Sigma(G; X) \text{ is the minimal possible value of } n_1 + \cdots + n_m$$

taken over all nonnegative integers $n_1, \dots, n_m$ satisfying

$$n_1\boldsymbol{v}_1 + \cdots + n_m\boldsymbol{v}_m = X\mathbf{1}. \tag{3}$$

It will be helpful to renormalize. Setting $\nu_i = n_i/X$, we seek to choose the n_i to minimize $\sum_{i=1}^m \nu_i$ subject to

$$\nu_1\boldsymbol{v}_1 + \cdots + \nu_m\boldsymbol{v}_m = \mathbf{1}. \tag{4}$$

We temporarily forget about the n_i and pretend that the problem, from the get-go, was to minimize the sum of nonnegative real variables $\nu_1, \dots, \nu_m$ subject only to the constraint (4). This is a linear

programming problem. Since the $\boldsymbol{v}_i$ are nonzero vectors with all entries nonnegative, the feasible polytope is bounded, and the minimum of $\sum_{i=1}^m \nu_i$ is attained at a vertex. Call this minimum σ. Since the $\boldsymbol{v}_i$ have rational entries (as does $\mathbf{1}$), all the vertices of the feasible polytope are rational, and hence so is σ. Thus, Proposition 1.1 will follow if we prove that $\operatorname{Clr} G = \sigma$.

If $n_1, \dots, n_m$ are nonnegative integers satisfying (3), our work in the last paragraph implies that $n_1/X + \cdots + n_m/X \geq \sigma$. Hence, $n_1 + \cdots + n_m \geq \sigma X$. We conclude that

$$\Sigma(G; X) \geq \sigma X. \tag{5}$$

To obtain a corresponding upper bound, let $\nu_1, \dots, \nu_m$ be nonnegative real numbers satisfying (4) for which $\sum_{i=1}^m \nu_i = \sigma$. Given a large even number X, we put $n_i' := \lfloor X\nu_i \rfloor$ for $i = 1, \dots, m$. After performing n_i' copies of the move $\boldsymbol{v}_i$, for $i = 1, \dots, m$, the number of leftover chips in each stack is tallied by the components of the vector

$$\begin{aligned} X\mathbf{1} - n_1'\boldsymbol{v}_1 - \cdots - n_m'\boldsymbol{v}_m &= X\mathbf{1} - ((X\nu_1 - \{X\nu_1\})\boldsymbol{v}_1 + \cdots + (X\nu_m - \{X\nu_m\})\boldsymbol{v}_m) \\ &= \{X\nu_1\}\boldsymbol{v}_1 + \cdots + \{X\nu_m\}\boldsymbol{v}_m, \end{aligned}$$

where $\{t\} := t - \lfloor t \rfloor$ denotes the fractional part of the real number t. Each $\{X\nu_i\} < 1$, and so the total number of leftover chips is bounded by the total number of chips involved in the moves $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ — a quantity independent of X. Any bounded number of chips can be swept up in a bounded number of moves. Hence, for some constant C (depending on G but not on X),

$$\Sigma(G; X) \leq n_1' + \cdots + n_m' + C \leq X(\nu_1 + \cdots + \nu_m) + C = \sigma X + C. \tag{6}$$

Comparing (5) and (6), we conclude that $\operatorname{Clr} G = \sigma$.

2.2. **Sketch of the proof of Theorem B.** To avoid becoming entangled in analytic unpleasantries, we will be intentionally vague in this section, referring the reader to [13] or [1] for the gory details. Our objective is to get to the meat of the proof of Theorem B, without choking on the bone.

Any study of elasticity must begin by recalling the origin story of irreducible factorizations. Let α be a nonzero, nonunit of $\mathcal{O}_K$, and factor (α) into (not necessarily distinct) prime ideals, say

$$(\alpha) = P_1 \cdots P_g.$$

Any factorization of α into irreducible elements of $\mathcal{O}_K$, say

$$\alpha = \pi_1 \cdots \pi_k,$$

corresponds to a partition of the multiset $\{P_1, \dots, P_g\}$ into multisets $\Pi_1, \dots, \Pi_k$, where the ideals in each Π_i multiply to (π_i). We observe that...

(a) The (classes of the) ideals in Π_i multiply to the identity in $\operatorname{Cl}(K)$.

(b) No proper subproduct of the ideals in Π_i comes out to the identity in $\operatorname{Cl}(K)$. (Otherwise, the complementary subproduct does as well. These two products are principal ideals, and their generators multiply — after adjusting by a unit — to the allegedly irreducible element π_i.)

Conversely, if $\Pi_1', \dots, \Pi_\ell'$ is any partition of the multiset $\{P_1, \dots, P_g\}$ satisfying conditions (a) and (b), then α has a factorization into irreducibles of length ℓ. Indeed, (a) allows us to write the product of the ideals in Π_i' as (π_i') for some nonunit $\pi_i' \in \mathcal{O}_K$ while (b) ensures that each π_i' is

irreducible. Finally, $(\pi_1' \cdots \pi_\ell') = P_1 \cdots P_g = (\alpha)$, so that after multiplying π_1' (say) by a suitable unit, we have $\pi_1' \cdots \pi_\ell' = \alpha$.

The upshot is a concrete description of the length spectrum $\mathcal{L}(\alpha)$ in terms of an α-dependent variant of $\mathrm{Cl}(K)$-solitaire. For each $C \in \mathrm{Cl}(K)$, we now take the number of chips in the C-pile as the number of prime ideals from the class C appearing in the factorization of α (with repeated prime ideal factors counted multiple times). Then $\mathcal{L}(\alpha)$ is the set of all k for which the table can be cleared in k moves.

So something *like* $\mathrm{Cl}(K)$-solitaire is in the picture. But why does the clearing constant $\operatorname{Clr}\mathrm{Cl}(K)$ — which came to us from analyzing *equal-height* games — come into play? To explain this, we need a bit of analytic number theory.

According to a 1917 theorem of G. H. Hardy and S. Ramanujan [7], asymptotically 100% of natural numbers n have "about" $\log\log n$ prime factors. (Here and below, primes and prime ideals are counted with multiplicity.) More precisely: For each $\epsilon > 0$, the proportion of $n \in (1, N]$ whose count of prime factors lies between $(1-\epsilon)\log\log n$ and $(1+\epsilon)\log\log n$ tends to 100%, as $N \to \infty$.

Similar results can be proved in the number field case. For each (fixed) number field K, almost all principal ideals (α) of $\mathcal{O}_K$ have about $\log\log|N\alpha|$ prime ideal factors. In fact, letting $h := \#\mathrm{Cl}(K)$ be the **class number of** K, almost all (α) have about $\frac{1}{h}\log\log|N\alpha|$ prime factors from each of the h ideal classes. The provenance of this last result is harder to trace; a recent reference is [14] (see Theorem 7 there for a stronger result).

Piecing together the algebra and the analysis: For almost all principal ideals (α), the length spectrum of α is determined by a variant of $\mathrm{Cl}(K)$-solitaire where the stacks have *almost* equal height (each $\approx \frac{1}{h}\log\log|N\alpha|$). This is starting to sound familiar!

When the stacks have precisely the same large (even) height X, we have seen that the minimum number of clearing moves is asymptotically $(\operatorname{Clr}\mathrm{Cl}(K))X$, as X grows. We have not yet said anything about the maximum number of moves, but this is comparatively straightforward to analyze. As usual, the 0-stack is cleared in X moves. Every other move involves at least two chips, which upper bounds the number of moves by $X + \frac{1}{2}(h-1)X = \frac{h+1}{2}X$. And this upper bound is easily achieved: For each $C \in \mathrm{Cl}(K)$, pair the stacks corresponding to C and its inverse. Remove one chip at a time from both — if the stack is its own inverse, remove two at a time. This clears the nonzero stacks in precisely $\frac{h-1}{2}X$ moves. So in this idealized X-chips/stack scenario, the ratio of the maximum number of possible moves to the minimum number is asymptotically

$$\frac{(h+1)X/2}{(\operatorname{Clr}\mathrm{Cl}(K))X} = \frac{1+\#\mathrm{Cl}(K)}{2\operatorname{Clr}\mathrm{Cl}(K)},$$

matching Theorem B.

To pass from the idealized back to the actual, we apply the reasoning of the last paragraph with two different values of X: First, the largest even X not exceeding the height of any stack, and second, the smallest even X at least the height of every stack. Both values of X are almost always $\approx \frac{1}{h}\log\log|N\alpha|$, and so almost always asymptotically equivalent. We conclude (with a little work) that for almost all (α), we have $\rho(\alpha) = \frac{\sup\mathcal{L}(\alpha)}{\inf\mathcal{L}(\alpha)} \approx \frac{1+\#\mathrm{Cl}(K)}{2\operatorname{Clr}\mathrm{Cl}(K)}$, as asserted in Theorem B.

3. Charging ahead: The cross number

If D is an integral domain, call $\pi \in D$ **primary** if π generates a primary ideal of D. Equivalently, π is a nonunit and for all $x, y \in D$,

$$\pi \mid xy \quad \text{implies} \quad \pi \mid x \text{ or } \pi \mid y^n \text{ for some positive integer } n.$$

In 1984, Krause gave the following characterization of number fields whose class groups are cyclic of prime power order [9].

Theorem C. *For each number field K, the following are equivalent:*

(i) $\mathrm{Cl}(K)$ *is cyclic of prime power order,*

(ii) *there is a positive integer m such that, for all irreducibles π of $\mathcal{O}_K$, the element π^m factors as a product of at most m primary elements.*

When (ii) *holds, the smallest possible value of m is precisely $\#\mathrm{Cl}(K)$.*

To prove Theorem C, Krause associates to each finite abelian group G an invariant he calls the "cross number." For each $\boldsymbol{v} \in \mathbb{R}^G$, define the **charge** of $\boldsymbol{v}$ as the real number

$$\sum_{g \in G} \frac{\boldsymbol{v}[g]}{\mathrm{ord}(g)},$$

where $\mathrm{ord}(g)$ denotes the order of g in G. Then the **cross number** of G, denoted $k(G)$, is the maximum charge of a legal move in G-solitaire.

Removing a chip from the 0-stack is a move of charge 1, and so we always have $k(G) \geq 1$. The "heavy lifting" in the proof of Theorem C consists in determining when equality holds.

Proposition 3.1 (see [9, Lemma 2])**.** *Let G be a finite abelian group of order larger than* 1*. Then G is cyclic of prime power order if and only if $k(G) = 1$.*

Rather than rehash Krause's proof of [9] of Theorem C, it will be more germane to call out how Proposition 3.1 immediately yields a nontrivial result about clearing constants.

Set up G-solitaire with X chips/pile. The heights of the piles in this initial configuration are recorded by the vector $X\mathbf{1}$, which has charge

$$X \sum_{g \in G} \frac{1}{\mathrm{ord}(g)}. \tag{7}$$

By Proposition 3.1, when G is cyclic of prime power order, each move depletes the charge of the configuration by at most 1. So to bring the charge down to 0, the number of moves must be at least (7). That is,

$$\Sigma(G; X) \geq X \sum_{g \in G} \frac{1}{\mathrm{ord}(g)}.$$

On the other hand, for any group G, the simple strategy of picking off the elements in the g-stack $\mathrm{ord}(g)$-at a time yields an essentially matching upper bound for $\Sigma(G; X)$, namely

$$\Sigma(G; X) \leq X \sum_{g \in G} \frac{1}{\mathrm{ord}(g)} + C,$$

where C is a constant depending on G but not on X. (The C comes from the fact that X may not be divisible by every $\mathrm{ord}(g)$. In that case, we will need to sweep up a bounded number of chips at the end.) It follows that

$$\mathrm{Clr}\,\mathbb{Z}/p^r\mathbb{Z} = \sum_{g\in\mathbb{Z}/p^r\mathbb{Z}} \frac{1}{\mathrm{ord}(g)} = 1 + \sum_{j=1}^{r} \frac{1}{p^j}\cdot(p^{j-1}(p-1)) = 1 + r\frac{p-1}{p}. \tag{8}$$

This simple proof is also a "proof of concept," suggesting that a refined understanding of $k(G)$, for general G, might have important implications for the study of clearing constants. Unfortunately, $k(G)$ is not so easy to get one's hands on!

There is at least a fairly straightforward lower bound for $k(G)$, noted already in Krause's initial paper [9]: Write G as a direct sum of cyclic groups of prime power order, say $G = \oplus_{i=1}^{k}\mathbb{Z}/p_i^{e_i}\mathbb{Z}$. Then

$$k(G) \geq \left(\sum_{i=1}^{k} \frac{p_i^{e_i}-1}{p_i^{e_i}}\right) + \frac{1}{\mathrm{Exp}\,G}, \tag{9}$$

where $\mathrm{Exp}\,G$ is the exponent of G, the least common multiple of the $p_i^{e_i}$. To see (9), let $\boldsymbol{v}$ be the move involving $p_1^{e_1}-1$ copies of $(1,0,\dots,0)$, $p_2^{e_2}-1$ copies of $(0,1,\dots,0)$, continuing through $p_k^{e_k}-1$ copies of $(0,0,\dots,1)$, and finishing with a single copy of $(1,1,\dots,1)$. Then the charge of $\boldsymbol{v}$ is precisely the right-hand side of (9).

Perhaps surprisingly, in every example we can compute, (9) is an exact equality. It is an open problem to decide whether or not equality holds universally. We *do* know equality for several classes of groups, including ...

- all abelian p-groups (A. Geroldinger [5]),
- all groups $\mathbb{Z}/pq\mathbb{Z}$ and $\mathbb{Z}/pqr\mathbb{Z}$ with p,q,r distinct primes (U. Krause and C. Zahlten; see [10, Theorems 3 and 4]),
- all groups $\mathbb{Z}/p^m\mathbb{Z}\oplus\mathbb{Z}/p^n\mathbb{Z}\oplus(\mathbb{Z}/q\mathbb{Z})^s$, with p,q distinct primes and m,n,s positive integers (A. Geroldinger and R. Schneider; see [6, Theorem 1]).

In particular, we have equality in (9) for $G = \mathbb{Z}/p\mathbb{Z}\oplus\mathbb{Z}/p^2\mathbb{Z}$, which will be used in our proof of Theorem 1.2 below.

4. The clearing constant of $\mathbb{Z}/p\mathbb{Z}\oplus\mathbb{Z}/p^2\mathbb{Z}$

4.1. **Economy of movement.** For every finite abelian group G, there is an easy lower bound on $\mathrm{Clr}\,G$ in terms of $k(G)$: To clear the table, the X chips in the 0-stack must be removed one-by-one, while the remaining chips have charge $X\sum_{g\in G, g\neq 0} 1/\mathrm{ord}(g)$. Therefore, $\Sigma(G;X) \geq X + \frac{X}{k(G)}\sum_{g\in G,\, g\neq 0} 1/\mathrm{ord}(g)$, and

$$\mathrm{Clr}\,G \geq 1 + \frac{1}{k(G)} \sum_{g\in G,\, g\neq 0} \frac{1}{\mathrm{ord}(g)}. \tag{10}$$

When equality holds in (10), we call the group G **economical**. For example, we saw in the previous section that the groups $\mathbb{Z}/p^r\mathbb{Z}$ are all economical. Allen and Pleasants proved, more

generally, that each group of the form $(\mathbb{Z}/p^r\mathbb{Z})^s$ is economical (this is implicit in [1, Theorem 5]). Equivalently (by a short computation),

$$\operatorname{Clr}(\mathbb{Z}/p^r\mathbb{Z})^s = 1 + \frac{p^{rs}(1-p^{-s})}{s(p^r-1)+1} \cdot \frac{1-p^{-(s-1)r}}{1-p^{-(s-1)}}, \tag{11}$$

where we take $\frac{1-p^{-(s-1)r}}{1-p^{-(s-1)}} = r$ when $s = 1$.

We will prove our Theorem 1.2 by showing that each $G = \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z}$ is also economical. By the results of §3,

$$k(G) = \frac{p-1}{p} + \frac{p^2-1}{p^2} + \frac{1}{p^2} = \frac{2p-1}{p}.$$

Since G contains p^2-1 elements of order p and p^3-p^2 elements of order p^2, we have that

$$\begin{aligned}\frac{1}{k(G)} \sum_{g\in G,\, g\neq 0} \frac{1}{\operatorname{ord}(g)} &= \frac{p}{2p-1}\left(\frac{p^2-1}{p} + \frac{p^3-p^2}{p^2}\right) \\ &= \frac{2p^2-p-1}{2p-1}.\end{aligned}$$

Thus, the right-hand side of (10) matches with the value for $\operatorname{Clr} G$ claimed in Theorem 1.2. So proving G economical will prove Theorem 1.2.

4.2. **Do you know it when you see it?** In this section, we develop a criterion allowing us to show efficiently that groups are economical (Proposition 4.2).

Recall that $\mathbb{L}_G$ denotes the set of moves in G-solitaire, thought of as a subset of $\mathbb{R}^G$. We let $\mathbb{L}_G^{\max} \subseteq \mathbb{L}_G$ denote the set of moves of maximum charge $k(G)$. We define $\boldsymbol{\chi} \in \mathbb{R}^G$ by $\boldsymbol{\chi}[0] = 0$ and $\boldsymbol{\chi}[g] = 1$ for all $g \in G, g \neq 0$. The following is "version 0" of our criterion.

Lemma 4.1. *G is economical if $\boldsymbol{\chi}$ belongs to the nonnegative span of $\mathbb{L}_G^{\max}$.*

Proof. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ be a list of the moves in $\mathbb{L}_G^{\max}$, and suppose that

$$\boldsymbol{\chi} = c_1\boldsymbol{v}_1 + \cdots + c_m\boldsymbol{v}_m. \tag{12}$$

Starting from stacks of size X in G-solitaire, clear the X chips in the 0-stack first. Then perform the move $\boldsymbol{v}_1$ $\lfloor c_1 X \rfloor$ times, the move $\boldsymbol{v}_2$ $\lfloor c_2 X \rfloor$ times, etc. This leaves only a bounded number of chips, which can be dealt with in a bounded number of moves (compare with the arguments of §2.1). Hence, for some constant C depending only on G,

$$\Sigma(G; X) \leq X + (c_1 + \cdots + c_m)X + C. \tag{13}$$

Comparing the charge on the two sides of (12), we find that

$$\sum_{g\in G,\, g\neq 0} \frac{1}{\operatorname{ord}(g)} = k(G)(c_1 + \cdots + c_m).$$

Hence,

$$c_1 + \cdots + c_m = \frac{1}{k(G)} \sum_{g\in G, g\neq 0} \frac{1}{\operatorname{ord}(g)}.$$

Substituting this expression for $c_1 + \cdots + c_m$ into (13), dividing by X and sending $X \to \infty$ shows that

$$\operatorname{Clr} G \le 1 + \frac{1}{k(G)} \sum_{g \in G,\, g \neq 0} \frac{1}{\operatorname{ord}(g)}.$$

However, (10) tells us that the left-hand side is always at least as large as the right. So it must be that equality holds, proving G is economical. □

Lemma 4.1 is generally too cumbersome to apply directly, in part because the vectors involved live in a space of dimension equal to the order of G, which can be quite large. To cut down the dimension, we bring in the action of the automorphism group $\operatorname{Aut}(G)$ of G.

Say that $g, g' \in G$ are **automorphism-equivalent** if there is a $\sigma \in \operatorname{Aut}(G)$ with $\sigma(g) = g'$. We define $\tilde{G}$ as the quotient set corresponding to this equivalence relation. We refer to the elements of $\tilde{G}$ as **automorphism types**, denoting the type of $g \in G$ by $\tilde{g}$. For each $v \in \mathbb{R}^G$, the **profile** of $\boldsymbol{v}$ is the vector $\tilde{\boldsymbol{v}} \in \mathbb{R}^{\tilde{G}}$ with

$$\tilde{\boldsymbol{v}}[\tilde{g}] = \sum_{g' \in G,\, g' \sim g} \boldsymbol{v}[g'] \quad \text{for all } g \in G.$$

In words, $\tilde{\boldsymbol{v}}$ is determined from $\boldsymbol{v}$ by "grouping together contributions from automorphism-equivalent elements." We let $\widetilde{\mathbb{L}_G^{\max}}$ denote the collection of profiles of moves of maximum charge. Our workhorse criterion is the "tilde-ed" version of Lemma 4.1 (cf. [1, pp. 76–77]).

Proposition 4.2. *G is economical if $\tilde{\boldsymbol{\chi}}$ belongs to the nonnegative span of $\widetilde{\mathbb{L}_G^{\max}}$.*

Before giving the proof, we illustrate Proposition 4.2 with a toy example, proving that $G = (\mathbb{Z}/p\mathbb{Z})^s$ is economical for every choice of prime p and positive integer s. (This is the special case $r = 1$ of the earlier-quoted result of Allen–Pleasants.) Every two nonzero elements of the group $(\mathbb{Z}/p\mathbb{Z})^s$ are automorphism-equivalent, as any two nonzero vectors in the $\mathbb{Z}/p\mathbb{Z}$-vector space $(\mathbb{Z}/p\mathbb{Z})^s$ can be interchanged by an invertible linear transformation. So there is precisely one nontrivial automorphism class, and $\tilde{G} \cong \mathbb{Z}/2\mathbb{Z}$. We identify $\mathbb{R}^{\tilde{G}}$ with $\mathbb{R}^2$, with the nontrivial automorphism type indexing the first component. Then $\tilde{\boldsymbol{\chi}} = \big[p^s - 1, 0\big]$. Now let $\boldsymbol{v}$ be the move in G-solitaire involving $p-1$ copies of each of $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 1)$, and one copy of $(1, 1, \ldots, 1)$. Then $\tilde{\boldsymbol{v}} = \big[s(p-1) + 1, 0\big]$. Clearly, $\tilde{\boldsymbol{\chi}}$ is a positive real multiple of $\tilde{\boldsymbol{v}}$. Hence, G is economical by Proposition 4.2.

We require one more piece of notation before explaining the proof of Proposition 4.2. If $\boldsymbol{v} \in \mathbb{R}^G$ and $\sigma \in \operatorname{Aut}(G)$, we define $\boldsymbol{v}^\sigma \in \mathbb{R}^G$ by

$$\boldsymbol{v}^\sigma[g] = \boldsymbol{v}[\sigma(g)] \quad \text{for all } g \in G.$$

Proof of Proposition 4.2. By Lemma 4.1, it suffices to show that if $\tilde{\boldsymbol{\chi}}$ lies in the nonnegative span of $\widetilde{\mathbb{L}_G^{\max}}$, then $\boldsymbol{\chi}$ belongs to the nonnegative span of $\mathbb{L}_G^{\max}$. As before, let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ be a list of the moves in $\mathbb{L}_G^{\max}$. Suppose there are nonnegative $c_1, \ldots, c_m$ with

$$c_1 \tilde{\boldsymbol{v}_1} + \cdots + c_m \tilde{\boldsymbol{v}_m} = \tilde{\boldsymbol{\chi}}.$$

We claim that

$$\sum_\sigma (c_1 \boldsymbol{v}_1^\sigma + \cdots + c_m \boldsymbol{v}_m^\sigma) = (\#\operatorname{Aut}(G))\boldsymbol{\chi}, \tag{14}$$

where the sum is over all $\sigma \in \mathrm{Aut}(G)$. Since $\boldsymbol{v}$ and $\boldsymbol{v}^\sigma$ share the same charge (for each $\boldsymbol{v} \in \mathbb{R}^G$ and $\sigma \in \mathrm{Aut}(G)$), dividing (14) through by $\#\mathrm{Aut}(G)$ exhibits $\boldsymbol{\chi}$ as a nonnegative combination of elements of $\mathbb{L}_G^{\max}$.

The 0-components of the left and right-hand sides of (14) both vanish. Now let $g \neq 0$. The gth component of the left of (14) is given by

$$\begin{aligned}\sum_\sigma (c_1 \boldsymbol{v}_1^\sigma[g] + \cdots + c_m \boldsymbol{v}_m^\sigma[g]) &= \sum_\sigma (c_1 \boldsymbol{v}_1[\sigma(g)] + \cdots + c_m \boldsymbol{v}_m[\sigma(g)]) \\ &= \sum_\sigma (c_1 \boldsymbol{v}_1 + \cdots + c_m \boldsymbol{v}_m)[\sigma(g)].\end{aligned}$$

As σ runs through $\mathrm{Aut}(G)$, the elements $\sigma(g)$ run $\#\mathrm{Aut}(G)/\#\mathrm{Orb}(g)$ times through the automorphism orbit $\mathrm{Orb}(g)$ of g. Also, $\tilde{\boldsymbol{\chi}}[\tilde{g}] = \sum_{g' \sim g} 1 = \#\mathrm{Orb}(g)$. Therefore,

$$\begin{aligned}\sum_\sigma (c_1 \boldsymbol{v}_1 + \cdots + c_m \boldsymbol{v}_m)[\sigma(g)] &= \widetilde{(c_1 \boldsymbol{v}_1 + \cdots + c_m \boldsymbol{v}_m)}[\tilde{g}] \cdot \frac{\#\mathrm{Aut}(G)}{\#\mathrm{Orb}(g)} \\ &= \tilde{\boldsymbol{\chi}}(\tilde{g}) \cdot \frac{\#\mathrm{Aut}(G)}{\#\mathrm{Orb}(g)} \\ &= \#\mathrm{Orb}(g) \cdot \frac{\#\mathrm{Aut}(G)}{\#\mathrm{Orb}(g)} \\ &= \#\mathrm{Aut}(G),\end{aligned}$$

which agrees with the gth component on the right of (14). □

4.3. **Application to the proof of Theorem 1.2.** Throughout this section, $G = \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z}$ for a prime p. The following lemma determines the automorphism orbits of G.

Lemma 4.3. *Let p be a prime. There are three automorphism orbits on G:*

(i) *elements of order p of the form $(0, pv)$, where $v \not\equiv 0 \pmod p$,*

(ii) *elements of order p of the form (u, pv), where $u \not\equiv 0 \pmod p$, v is arbitrary,*

(iii) *elements of order p^2, i.e., of the form (u, v) with u arbitrary, $v \not\equiv 0 \pmod p$.*

Proof. It is straightforward to check that each nonzero element of G is described by exactly one of (i)–(iii). Next, order considerations show that no element in (i) or (ii) is carried by an automorphism to one described by (iii). Furthermore, nothing in (i) is automorphism-equivalent to something in (ii): The elements in (i) belong to pG while those in (ii) do not, and belonging to pG is preserved under automorphism.

It remains to show that for each of (i)–(iii), all elements of that form are automorphism-equivalent. This is easy for (i): If $v, v' \not\equiv 0 \pmod p$, there is an integer $r \not\equiv 0 \pmod p$ with $rv \equiv v' \pmod p$. "Multiply by r in the second component" is an automorphism of G carrying $(0, pv)$ to $(0, pv')$. (Here we have noted that $rv \equiv v' \pmod p$ implies $rpv \equiv pv' \pmod{p^2}$.)

We turn now to (ii): If $u \not\equiv 0 \pmod p$, we can choose $r \in \mathbb{Z}$ with $ru \equiv 1 \pmod p$. Then "multiply by r in the first component" takes (u, pv) to $(1, pv)$. If $v \equiv 0 \pmod p$, then $(1, pv) = (1, 0)$. If $v \not\equiv 0 \pmod p$, we can choose $r' \in \mathbb{Z}$ with $r'v \equiv 1 \pmod p$; in that case, "multiply by r' in the second component" takes $(1, pv)$ to $(1, p)$. Hence: To handle (ii), it suffices to show that $(1, 0)$ and $(1, p)$

are equivalent. An explicit automorphism taking $(1,0)$ to $(1,p)$ is given by $(x,y) \to (x, px+y)$. (Note that x is well-defined mod p, so that px is well-defined mod p^2.)

Finally, we deal with (iii). Scaling the second component will take any element in (iii) to the form $(u,1)$. Now scaling in the first component, we see that any two elements of the form $(u,1)$ with $u \not\equiv 0 \pmod{p}$ are automorphism-equivalent. This reduces the problem to showing that $(0,1)$ and $(1,1)$ are equivalent. But the map $(x,y) \mapsto (x-y,y)$ is an automorphism of G carrying $(1,1)$ to $(0,1)$. (Here $y \in \mathbb{Z}/p^2\mathbb{Z}$ is interpreted in $\mathbb{Z}/p\mathbb{Z}$ by reduction mod p, which is well-defined as p divides p^2.) □

Remark. Our proof of Lemma 4.3 is self-contained but ad hoc. One can also give a more conceptual argument, based on an explicit description of the automorphisms of G (see, e.g., [8]).

We now identify $\mathbb{R}^{\tilde{G}}$ with $\mathbb{R}^4$, indexing the first three components by automorphism types (i)—(iii), respectively, and indexing the fourth component by the type of 0. Straightforward counting arguments show that

$$\tilde{\boldsymbol{\chi}} = [p-1, p^2-p, p^3-p^2, 0].$$

Theorem 1.2 is now within easy reach.

Proof of Theorem 1.2. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ be the following G-solitaire moves:

$\boldsymbol{v}_1$: discard $(1,0)$ $p-1$ times, $(0,p)$ $p-1$ times, $(0,1)$ $p-1$ times, $(1,1)$ once,

$\boldsymbol{v}_2$: discard $(1,0)$ $p-1$ times, $(-1,p)$ $p-1$ times, $(0,p)$ once,

$\boldsymbol{v}_3$: discard $(1,0)$ $p-1$ times, $(0,1)$ p^2-1 times, $(1,1)$ once.

These are indeed legal moves (a straightforward check), and each has charge $2-\frac{1}{p}$, which we know from §3 to be the cross number of G. Moreover,

$$\tilde{\boldsymbol{v}}_1 = [p-1, p-1, p, 0],$$
$$\tilde{\boldsymbol{v}}_2 = [1, 2(p-1), 0, 0], \quad \textit{and}$$
$$\tilde{\boldsymbol{v}}_3 = [0, p-1, p^2, 0].$$

Solving the corresponding system of three equations in three variables, we find that $c_1\tilde{\boldsymbol{v}}_1 + c_2\tilde{\boldsymbol{v}}_2 + c_3\tilde{\boldsymbol{v}}_3 = \tilde{\boldsymbol{\chi}}$ for

$$c_1 = \frac{2p^2-3p}{2p^2-3p+1},$$
$$c_2 = \frac{1}{2p-1},$$
$$c_3 = \frac{2p^3-5p^2+2p+2}{2p^2-3p+1}.$$

As $p \geq 2$, all of $c_1, c_2, c_3 > 0$, and so the criterion of Proposition 4.2 is satisfied. □

Remark. Since we are primarily interested in positive results in this paper, we have not bothered about the converses of Lemma 4.1 and Proposition 4.2. But it is not so hard to show that those also hold. Here is an interesting application of this: One can verify computationally that $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2^3\mathbb{Z}$ fails the criterion of Proposition 4.2. Thus, $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2^3\mathbb{Z}$ is not economical! In particular, the above method of proof will not determine $\operatorname{Clr} \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^3\mathbb{Z}$.

It would seem an interesting problem to classify, even conjecturally, economical groups.

5. The clearing constant of $\mathbb{Z}/pq\mathbb{Z}$

Throughout this section, p and q denote distinct primes, and $G = \mathbb{Z}/pq\mathbb{Z}$. We prove Theorem 1.3 in two parts, by establishing corresponding upper and lower bounds on $\operatorname{Clr} G$. We start with the upper bound argument, which is pleasingly explicit.

Proof that $\operatorname{Clr} G \leq 3 - \frac{1}{p} - \frac{1}{q}$. For each $i = 1, 2, \ldots, p-1$ and $j = 1, 2, \ldots, q-1$, we introduce the move

$$\boldsymbol{v}_{i,j}: \quad \text{discard } iq \ p-1 \text{ times, } jp \ q-1 \text{ times, and } iq+jp \text{ once.}$$

Let's check that each $\boldsymbol{v}_{i,j}$ is a valid move. It is straightforward to compute that the removed chips sum to 0 — what we have to worry about is that a proper subcollection may also sum to 0. In that case, after perhaps swapping with the complementary subcollection, we can assume $iq + jp$ does not appear in the subcollection. Then iq appears a times and jp occurs b times, where a, b do not both vanish, $0 \leq a < p, 0 \leq b < q$, and

$$a(iq) + b(jp) \equiv 0 \pmod{pq}.$$

Looking at this equation modulo p, we see that p divides a. Similarly, looking mod q shows that q divides b. But now the ranges of a and b force $a = b = 0$, a contradiction.

Since $iq + jp$ is nonzero modulo both p and q, it represents an element of G of order pq. Therefore, each $\boldsymbol{v}_{i,j}$ has charge

$$\frac{p-1}{p} + \frac{q-1}{q} + \frac{1}{pq}.$$

We now clump together all $(p-1)(q-1)$ moves of the $\boldsymbol{v}_{i,j}$. When all these moves are performed one-after-another, we call it a run of the **Fundamental Macro**.

As i runs over the integers $0 < i < p$, the iq run over all the elements of order p in G, exactly once. Hence, a prescribed element of order p is removed $(p-1)(q-1)$ times by a single run of the Fundamental Macro. Similarly, any given element of order q is removed $(p-1)(q-1)$ times.

As i, j run over the integers $0 < i < p$ and $0 < j < q$, the integers $iq + jp$ run over the elements of order pq in $\mathbb{Z}/pq\mathbb{Z}$, each exactly once. (Those who have taught elementary number theory recently may remember this argument as a way to count units in the ring $\mathbb{Z}/pq\mathbb{Z}$; see for instance the proof of Theorem 3.7 in [11].) Hence, any given element of order pq is removed precisely once in a run of the Fundamental Macro.

Now assume — temporarily! — that X is divisible by $pq(p-1)(q-1)$. Run the Fundamental Macro $\frac{X}{(p-1)(q-1)}$ times. This clears the stacks corresponding to elements of order p and elements of order q but leaves the 0-stack untouched and leaves $X - \frac{X}{(p-1)(q-1)}$ chips in each order pq stack. We clear each of the $(p-1)(q-1)$ order pq stacks pq chips at a time, and then clear the 0-stack one chip at a time. In total, the number of moves involved is

$$\underbrace{\frac{X}{(p-1)(q-1)} \cdot (p-1)(q-1)}_{\text{from running the Fun. Macro}} + \underbrace{\frac{X - X/(p-1)(q-1)}{pq} \cdot (p-1)(q-1)}_{\text{from the order } pq \text{ stacks}} + \underbrace{X}_{\text{0-stack}}.$$

Applying some elbow grease, this simplifies to

$$X\left(3-\frac{1}{p}-\frac{1}{q}\right).$$

In general, let X' be the largest multiple of $pq(p-1)(q-1)$ not exceeding X. Carrying out the procedure of the last paragraph, we clear the top X' elements of each stack in $X'(3-1/p-1/q) \le X(3-1/p-1/q)$ moves. Afterwards, only a bounded number of chips remain to be swept up, and hence $\Sigma(G;X) \le X(3-1/p-1/q)+C$, for some constant C. It follows that $\operatorname{Clr} G \le 3-\frac{1}{p}-\frac{1}{q}$, as claimed. □

Obtaining the corresponding lower bound on $\operatorname{Clr} G$ requires some preparation. The following fundamental result on Davenport constants was alluded to in the introduction.

Lemma 5.1. *Let H be an abelian group of order n. Then*

$$\operatorname{Dav} H \le n,$$

with equality if H is cyclic.

Proof. Let $h_1, h_2, \dots, h_n$ be any n-element sequence from H. Consider the $(n+1)$-element list

$$0, \quad h_1, \quad h_1+h_2, \quad \dots, \quad h_1+h_2+\cdots+h_n.$$

As $\#H = n$, two elements on this list must coincide. Thus, there are integers $0 \le i < j \le n$ with $h_1+\cdots+h_i = h_1+\cdots+h_j$. But then

$$h_{i+1}+\cdots+h_j = 0.$$

This proves that $\operatorname{Dav} H \le n$. If H is cyclic with generator h, then $h, h, h, \dots, h$ ($n-1$ times) is an $(n-1)$-element sequence in H with no zero-sum subsequence. Thus, $\operatorname{Dav} H$ cannot be smaller than n, forcing $\operatorname{Dav} H = n$. □

Lemma 5.2. *If $\boldsymbol{v} \in \mathbb{L}_G$ removes exactly i chips of order q, then $\boldsymbol{v}$ has charge at most*

$$1+\frac{p-1}{pq}i. \tag{15}$$

In particular, if no chips of order q are removed, then $\boldsymbol{v}$ has charge at most 1.

Proof. If $\boldsymbol{v}$ has charge at most 1, the claimed upper bound is obvious.

Thus, we may assume that every chip removed by $\boldsymbol{v}$ is nonzero. We may also assume that $i < q$. Indeed, suppose $i \ge q$, and let $g_1, \dots, g_i$ be the corresponding order q elements of G. Since G is cyclic of order pq, every element of order q belongs to the q-element subgroup $H = pG$ of G. By Lemma 5.1, there is a subsequence of $g_1, \dots, g_q$ summing to 0 in pG, and hence to 0 in G. Since $\boldsymbol{v}$ is a legal move, $i = q$, and the q chips of order q are the *only* chips removed by $\boldsymbol{v}$, so that $\boldsymbol{v}$ has charge $q/q = 1$.

Let's say $\boldsymbol{v}$ removes $i+j+k$ chips in total, where i is as in the lemma statement, and j and k count chips of orders p and pq, respectively. We list the corresponding elements of G as

$$a_1, \dots, a_i, \quad b_1, \dots, b_j, \quad c_1, \dots, c_k. \tag{16}$$

Then $\boldsymbol{v}$ has charge

$$\frac{i}{q}+\frac{j}{p}+\frac{k}{pq}. \tag{17}$$

The argument of the last paragraph, showing we may assume $i<q$, also lets us assume that $j<p$.

If the upper bound of the lemma fails, comparing (15) and (17) shows that

$$k > q(p-1-j)+q-i. \tag{18}$$

We assume (18) holds and arrive at contradiction.

Break $c_1,\dots,c_{q(p-1-j)}$ into $p-1-j$ consecutive blocks of length q. Applying Lemma 5.1 with $H=G/qG$, each of the $p-1-j$ blocks has a subsequence summing to an element of qG. We label these subsequence sums as $s_1,\dots,s_{p-1-j}$. A similar application of Lemma 5.1 shows that the q-term sequence

$$a_1,\dots,a_i,c_{q(p-1-j)+1},\dots,c_{q(p-1-j)+q-i}$$

has a subsequence whose sum, call it s, belongs to qG.

Now consider the sequence

$$b_1,\dots b_j,s_1,\dots,s_{p-1-j},s. \tag{19}$$

By construction, every subsequence sum of (19) is also a subsequence sum of

$$a_1,\dots,a_i,\quad b_1,\dots,b_j,\quad c_1,\dots,c_{q(p-1-j)+q-i}. \tag{20}$$

We call upon Lemma 5.1 one final time: The sequence (19) contains a subsequence summing to 0, since the p terms of (19) all belong to the p-element group qG. Hence, the $i+j+(q(p-1-j)+q-i)$-term sequence (20) also has 0 as a subsequence sum. But then the legality of $\boldsymbol{v}$ forces (20) to account for all of the chips removed by $\boldsymbol{v}$, contradicting (16) and (18). □

Proof that $\operatorname{Clr} G \geq 3-\frac{1}{p}-\frac{1}{q}$. It suffices to show that

$$\Sigma(G;X) \geq X\left(3-\frac{1}{p}-\frac{1}{q}\right) \tag{21}$$

for all (even) X.

Start with any sequence of moves clearing the stacks of height X. We may assume without loss of generality that the moves of charge larger than 1 are performed first. Each of those involves i elements of order q for some integer i with $0<i<q$. We let n_i denote the number of moves of charge larger than 1 involving precisely i elements of order q.

The charge of our initial configuration is

$$T := X\sum_{g\in G}\frac{1}{\operatorname{ord}(g)} = X\left(1+\frac{p-1}{p}+\frac{q-1}{q}+\frac{(p-1)(q-1)}{pq}\right). \tag{22}$$

After the $n_1+\cdots+n_{q-1}$ moves of charge larger than 1 are performed, the charge of the remaining configuration is, by Lemma 5.2, at least

$$T-\sum_{i=1}^{q-1}n_i\left(1+\frac{p-1}{pq}i\right).$$

All the remaining chips must be cleared using moves of charge at most 1, and so the last expression also serves as a lower bound on the number of remaining moves. Therefore,

$$\begin{aligned}\Sigma(G;X) &\geq \sum_{i=1}^{q-1} n_i + \left(T - \sum_{i=1}^{q-1} n_i \left(1 + \frac{p-1}{pq} i\right)\right) \\ &= T - \frac{p-1}{pq} \sum_{i=1}^{q-1} i n_i \\ &\geq T - \frac{p-1}{pq} \cdot X(q-1) \\ &= T - X\frac{(p-1)(q-1)}{pq}.\end{aligned}$$

(In going from the second line to the third, we use that $\sum_{i=1}^{q-1} i n_i$ does not exceed the total number of chips of order q, which is $X(q-1)$.) Referring back to our earlier expression (22) for T, we see that the final expression in the last display is precisely

$$X\left(1 + \frac{p-1}{p} + \frac{q-1}{q}\right) = X\left(3 - \frac{1}{p} - \frac{1}{q}\right).$$

This completes the proof of (21) as well as the proof of Theorem 1.3. □

Acknowledgements. We thank Matti Klock for her help with Figure 1. When we began work on this paper, P.P. was supported by NSF award DMS-2001581.

ChatGPT 5.2 was a helpful conversation partner throughout. In particular it played a crucial role in our proof that $G = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$ is *not* economical, by generating efficient `gp/PARI` code (human-checked for correctness!) efficiently enumerating all moves in G-solitaire.

References

[1] S. Allen and P. A. B. Pleasants, *The number of different lengths of irreducible factorization of a natural number in an algebraic number field*, Acta Arith. **36** (1980), 59–86.
[2] L. Carlitz, *A characterization of algebraic number fields with class number two*, Proc. Amer. Math. Soc. **11** (1960), 391–392.
[3] S. T. Chapman, *On the Davenport constant, the cross number, and their application in factorization theory*, Zero-dimensional commutative rings (Knoxville, TN, 1994), Lecture Notes in Pure and Appl. Math., vol. 171, Dekker, New York, 1995, pp. 167–190.
[4] S. T. Chapman and J. Coykendall, *Half-factorial domains, a survey*, Non-Noetherian commutative ring theory, Math. Appl., vol. 520, Kluwer Acad. Publ., Dordrecht, 2000, pp. 97–115.
[5] A. Geroldinger, *The cross number of finite abelian groups*, J. Number Theory **48** (1994), 219–223.
[6] A. Geroldinger and R. Schneider, *The cross number of finite abelian groups. II*, European J. Combin. **15** (1994), 399–405.
[7] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Math. **48** (1917), 76–92.
[8] C. J. Hillar and D. L. Rhea, *Automorphisms of finite abelian groups*, Amer. Math. Monthly **114** (2007), 917–923.
[9] U. Krause, *A characterization of algebraic number fields with cyclic class group of prime power order*, Math. Z. **186** (1984), 143–148.
[10] U. Krause and C. Zahlten, *Arithmetic in Krull monoids and the cross number of divisor class groups*, Mitt. Math. Ges. Hamburg **12** (1991), 681–696.
[11] W. J. LeVeque, *Elementary theory of numbers*, second ed., Dover Books on Advanced Mathematics, Dover Publications, Inc., New York, 1990.

[12] W. Narkiewicz, *A note on elasticity of factorizations*, J. Number Theory **51** (1995), 46–47.
[13] W. Narkiewicz and J. Śliwa, *Normal orders for certain functions associated with factorizations in number fields*, Colloq. Math. **38** (1977/78), 323–328.
[14] P. Pollack, *An elemental Erdős-Kac theorem for algebraic number fields*, Proc. Amer. Math. Soc. **145** (2017), 971–987.
[15] J.-L. Steffan, *Longueurs des décompositions en produits d'éléments irréductibles dans un anneau de Dedekind*, J. Algebra **102** (1986), 229–236.
[16] I. Stewart and D. Tall, *Algebraic number theory and Fermat's last theorem*, fifth ed., CRC Press, Boca Raton, FL, 2025.
[17] R. J. Valenza, *Elasticity of factorization in number fields*, J. Number Theory **36** (1990), 212–218.

Department of Mathematics, University of Georgia, Athens, GA 30602

Email address: pollack@uga.edu

Department of Mathematics and Computer Science, Lake Forest College, Lake Forest, IL 60045

Email address: trevino@lakeforest.edu